

I created a knowledge base as a simple pickle file of the term frequency dictionary created in my program. Some of the top 40 terms from the corpus are shown below.

```
[('ranger', 0.025928349419270406),  
 ('texas', 0.025764763618517914),  
 ('team', 0.008342875838377228),  
 ('name', 0.006952396531981024),  
 ('baseball', 0.006216260428594798),  
 ('history', 0.0055619172255848195),  
 ('blue', 0.0055619172255848195),  
 ('white', 0.004989366922951088),  
 ('red', 0.004907574022574841),  
 ('washington', 0.004662195321446098),  
 ('said', 0.004662195321446098),  
 ('one', 0.0042532308195648615),  
 ('al', 0.004171437919188614),  
 ('new', 0.003926059218059872),  
 ('west', 0.0037624734173073776),  
 ('first', 0.003598887616554883),  
 ('senator', 0.003598887616554883),  
 ('home', 0.0035170947161786356),  
 ('capture', 0.0034353018158023883),  
 ('short', 0.0034353018158023883),  
 ('state', 0.0031899231146736463),  
 ('american', 0.0030263373139211517),  
 ('page', 0.0029445444135449044),  
 ('year', 0.002862751513168657),  
 ('would', 0.002862751513168657),  
 ('mexican', 0.0026991657124161624),  
 ('logo', 0.002617372812039915),  
 ('violence', 0.002617372812039915),  
 ('life', 0.0025355799116636678),  
 ('jersey', 0.0025355799116636678),  
 ('time', 0.0025355799116636678),  
 ('u', 0.0025355799116636678),
```

Since the corpus was based off of the wikipedia article about the history of the Texas Rangers MLB team, the top 10 most important words should relate to the history of the team. Using my knowledge of baseball and the Rangers, I have chosen the following 10 words as the most important words

Thomas Bennett
CS 4395.001 - UTD Spring '23
Web Scraper Narrative

1. Ranger
2. Texas
3. Baseball
4. Washington
5. AL
6. West
7. Senator
8. American
9. Uniform
10. Field

The knowledge base is just the term frequency dictionary saved as a pickle file.

Sample Dialog

user> When was the team created?

bot> Which team?

user> The Texas Rangers

bot> The Texas Rangers first became a team in 1961, though they were known then as the Washington Senators. They changed their name to the Texas Rangers after moving to Arlington, Texas in 1972.